Introduction to artificial neural networks

Article *in* European Journal of Gastroenterology & Hepatology · January 2008

DOI: 10.1097/MEG.0b013e3282f198a0 · Source: PubMed

2 authors:

Enzo Grossi
Foundation Villa Santa Maria
732 PUBLICATIONS **11,376** CITATIONS
SEE PROFILE

Massimo Buscema
Semeion Centro Ricerche di Scienze della Comunicazione
324 PUBLICATIONS **5,565** CITATIONS
SEE PROFILE

Introduction to artificial neural networks

Enzo Grossi[a] and Massimo Buscema[b]

The coupling of computer science and theoretical bases such as nonlinear dynamics and chaos theory allows the creation of 'intelligent' agents, such as artificial neural networks (ANNs), able to adapt themselves dynamically to problems of high complexity. ANNs are able to reproduce the dynamic interaction of multiple factors simultaneously, allowing the study of complexity; they can also draw conclusions on individual basis and not as average trends. These tools can offer specific advantages with respect to classical statistical techniques. This article is designed to acquaint gastroenterologists with concepts and paradigms related to ANNs. The family of ANNs, when appropriately selected and used, permits the maximization of what can be derived from available data and from complex, dynamic, and multidimensional phenomena, which are often poorly predictable in the traditional 'cause and effect' philosophy. *Eur J Gastroenterol Hepatol* **19:1046–1054**

European Journal of Gastroenterology & Hepatology 2007, **19:**1046–1054

Keywords: artificial neural networks, diagnosis, evolutionary algorithms, nonlinearity, prognosis

[a]Bracco Spa Medical Department, Milan, Italy and [b]Semeion Research Centre for Science and Communication, Trigoria, Rome

Correspondence to Enzo Grossi, Medical Department, Bracco SpA, Via E. Folli 50 20136 Milano
E-mail: enzo.grossi@bracco.com

Received 29 August 2007 **Accepted** 29 August 2007

Introduction

The aim of this article is to discuss the possible advantages derived from the use of artificial neural networks (ANNs), which are some of the more advanced artificial intelligence tools today available, as an appropriate means of answering the emerging issues and 'demands' of medical science, and in particular of gastroenterology.

We are currently facing a paradox of sorts in which the amount of progress in the quality of the delivery of medical care in the everyday routine context of gastrointestinal diseases in the 'real world' is far from being proportional to the amount of scientific knowledge built up in basic science.

Different explanations can be given for this paradox. The increasing amount of clinical, laboratory, and diagnostic imaging information data requires more and more specific tools able to gather and recompose this information, and these tools are not easily available today as the traditional statistical reductionistic approach tends to 'see' things individually, to simplify, and to look at one single element at a time.

The increasing complexity of clinical data requires the use of mathematical models that are able to capture the key properties of the entire ensemble, including the linkages and hubs. The advancement of knowledge and the progress of understanding the nature of bodily rhythms and processes have shown that complexity and nonlinearity are ubiquitous in living organisms. These rhythms arise from stochastic, nonlinear biological mechanisms interacting with fluctuating environments. We need, for this reason, a special kind of mathematics that has, historically, remained cast away from the medical context.

In simple words, we have a problem of quantity and quality of medical information, which can be more appropriately addressed by the use of new computational tools such as ANNs.

What are artificial neural networks?

ANNs are artificial adaptive systems that are inspired by the functioning processes of the human brain [1]. They are systems that are able to modify their internal structure in relation to a function objective. They are particularly suited for solving problems of the nonlinear type, being able to reconstruct the fuzzy rules that govern the optimal solution for these problems.

The base elements of the ANN are the nodes, also called processing elements (PE), and the connections. Each node has its own input, from which it receives communications from other nodes and/or from the environment and its own output, from which it communicates with other nodes or with the environment. Finally, each node has a function *f* through which it transforms its own global input into output (Fig. 1). Each connection is characterized by the strength with which pairs of nodes are excited or inhibited. Positive values indicate excitatory connections, the negative ones inhibitory connections [2,3]. The connections between the nodes can modify themselves

0954-691X

Fig. 1

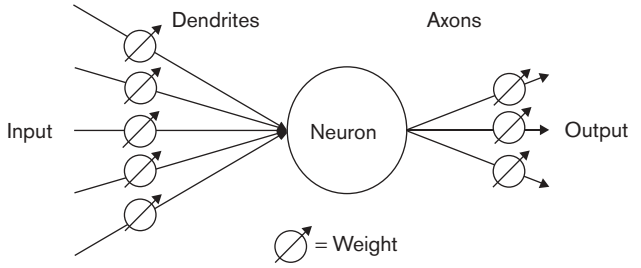
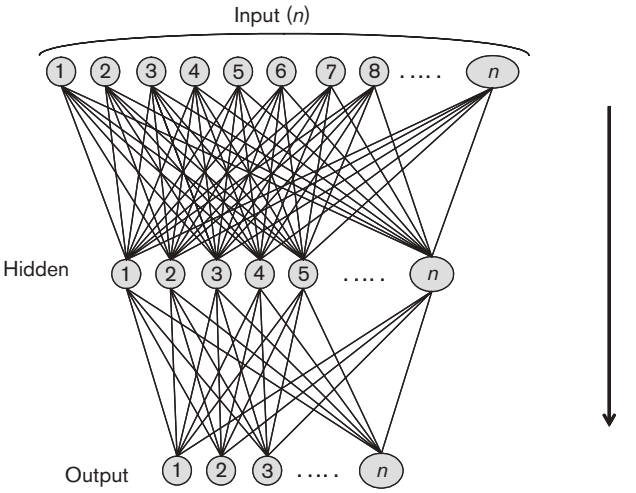


Diagram of a single processing element (PE) containing a neuron, weighted dendrites, and axons to process the input data and calculate an output.

Fig. 2



Back propagation neural network architecture.

over time. This dynamic starts a learning process in the entire ANN [4,5]. The way through which the nodes modify themselves is called 'Law of Learning'. The total dynamic of an ANN is tied to time. In fact, for the ANN to modify its own connections, the environment has to necessarily act on the ANN more times [6]. Data are the environment that acts on the ANN. The learning process is, therefore, one of the key mechanisms that characterize the ANN, which are considered adaptive processing systems. The learning process is one way to adapt the connections of an ANN to the data structure that make up the environment and, therefore, a way to 'understand' the environment and the relations that characterize it [7–10].

Neurons can be organized in any topological manner (e.g. one- or two-dimensional layers, three-dimensional blocks or more-dimensional structures), depending on the quality and amount of input data. The most common ANNs are composed in a so-called feed forward topology [11,12]. A certain number of PEs is combined to an input layer, normally depending on the amount of input variables. The information is forwarded to one or more hidden layers working within the ANN. The output layer, as the last element of this structure, provides the result. The output layer contains only one PE, whether the result is a binary value or a single number. Figure 2 represents the most popular architecture of neural networks: back propagation [13,14].

All PEs within the ANN are connected to other PEs in their neighbourhood. The way these connections are made might differ between the subtypes of neural networks [15,16]. Each of these connections has a so-called weight, which modifies the input or output value. The value of these connection weights is determined during the training process. This functionality is the basis for the ANN's learning capability. Therefore, it is important to understand that there are no classification rules written into the algorithm. The network just learns to understand and classify input patterns from examples.

Basic neural networks can normally be obtained with statistical computer software packages. Some companies offer specialized software to work with different neural networks (e.g. NeuralWorks Professional by NeuralWare Inc., Carnegie, Pennsylvania, USA [17] or CLEMENTINE Data Mining tool by Integral Solutions Limited, UK [18]). These software packages must be flexible and easy to handle for use in widespread purposes.

Properties of artificial neural networks

ANNs are high in pattern recognition-like abilities, which are needed for pattern recognition and decision-making and are robust classifiers with the ability to generalize and make decisions from large and somewhat fuzzy input data. ANNs are modeling mechanisms particularly skilled in solving nonlinear problems. In technical terms, we can say that a system is not complex when the function representing it is linear, that is when these two equations apply:

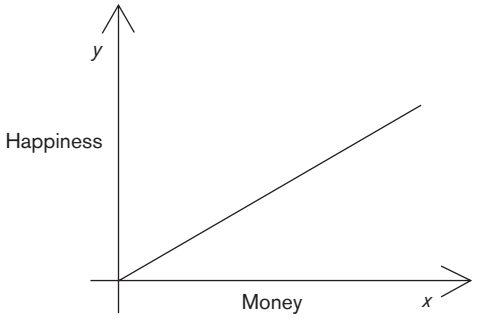
$$f(cx) = cf(x)$$

and

$$f(x_1 + x_2) = f(x_1) + f(x_2)$$

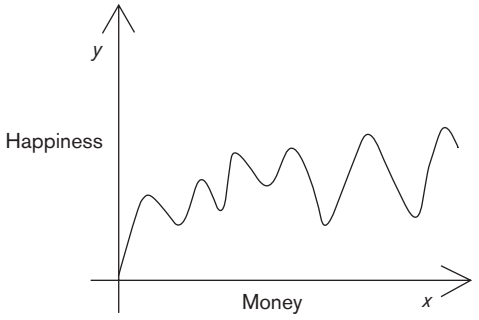
A complex, nonlinear system violates one or both of these conditions. Briefly, the more the function $y = f(x)$ is nonlinear, the more valuable it is to use an ANN to try and understand the rules, R, which govern the behavior inside the black box. If we take a Cartesian chart in which axis x represents the money a person gets, and axis y measures the degree of happiness that person obtains as a result, then:

Fig. 3



Linear relation between money and happiness.

Fig. 4



Nonlinear relation between money and happiness.

According to Fig. 3, the more money a person has, the happier he is. In this scenario, sampling of experimental data should offer the possibility of easily infering the relation and it would be worthless to use an ANN to analyze such a problem; unfortunately this scenario, like others in real life, is actually more an exception rather than a rule.

In real life, the relations are generally more complex, such as that presented in Fig. 4. In fact, as many of us can witness, an increase in earning can sometimes produce fears of losing money or uncertainties over how to invest this money, and this can reduce the feeling of happiness. The complex relation described in Fig. 4 does not permit us to understand, at first glance, from the data gathered experimentally, the relationship between money and happiness. In situations such as this, it makes sense to use an ANN to define precisely the relationship between money and happiness.

Artificial neural networks reveal the hidden rules of a problem

ANNs are data processing mechanisms that do not follow specific rules to process data, but which use the data they receive to discover the rules governing them [19,20]. This makes ANNs particularly useful in solving a problem for which we have the data involved, but do not know how those data are related to one another.

Artificial neural networks are adaptive and dynamically discover the fuzzy rules that connect various sets of data

This means that if they receive certain data in one phase, ANNs focus on certain rules; but if they later receive new and different data, ANNs will adjust their rules in accordance, integrating the old data with the new, and they can do this without any external instruction [21–24].

The continuous updating of data under their management creates a dynamic bank, whose rules are automatically refined by the ANNs as the problem in question evolves through time. This passage from an early categorization to a later, finer, and more complex one is managed by the ANN alone, using the new cases as data to learn about the new category.

Artificial neural networks can generalize, then predict and recognize

Once an ANN has been trained with suitable data to find the hidden rules governing a certain phenomenon, it is then able to correctly generalize to data it has never seen before (new, dirty, incomplete data, etc.).

When are they used?

The most typical problem that an ANN can deal with can be expressed as follows: given N variables, about which it is easy to gather data, and M variables, which differ from the first and about which it is difficult and costly to gather data, assess whether it is possible to predict the values of the M variables on the basis of the N variables.

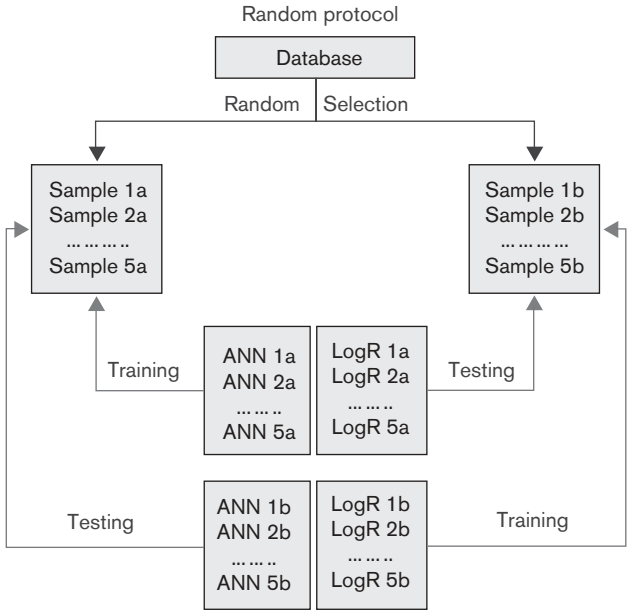
When the M variables occur subsequently in time to the N variables, the problem is described as a prediction problem; when the M variables depend on some sort of static and/or dynamic factor, the problem is described as one of recognition and/or discrimination and/or extraction of fundamental traits.

To correctly apply an ANN to this type of problem, we need to run a validation protocol. We must start with a good sample of cases, in each of which the N variables (known) and the M variables (to be discovered) are both known and reliable.

The sample of complete data is needed to:

- train the ANN, and
- assess its predictive performance.

Fig. 5



The 5 × 2 cross-validation protocol.

The validation protocol uses part of the sample to train the ANN (training set), whereas the remaining cases are used to assess the predictive capability of the ANN (testing set or validation set).

In this way we are able to test the reliability of the ANN in tackling the problem before putting it into operation.

Different types of protocol exist in literature, each presenting advantages and disadvantages. One of the most popular employed is the so-called 5×2 cross-validation protocol [25], which produces 10 elaborations for every sample. It consists in dividing the sample five times in two different subsamples, containing similar distribution of participants and controls (Fig. 5).

Description of the standard validation protocol

The protocol, from the point of view of a general procedure, consists of the following steps:

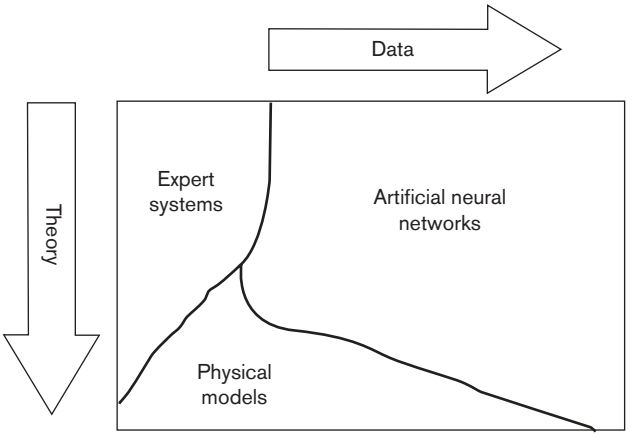
1. Subdividing the database in a random way into two subsamples: the first named training set and the second called testing set
2. Choosing a fixed ANN, and/or another model, which is trained on the training set, in this phase the ANN learns to associate the input variables with those that are indicated as targets
3. At the end of the training phase, the weight matrix produced by the ANN is saved and frozen together with all of the other parameters used for the training
4. With the weight matrix having saved the testing set, which it has not seen before, is shown to the ANN so that in each case the ANN can express an evaluation based on the previously carried out training, this operation takes place for each input vector and every result (output vector) is not communicated to the ANN
5. The ANN is in this way evaluated only in reference to the generalization ability that it has acquired during the training phase
6. A new ANN is constructed with identical architecture to the previous one and the procedure is repeated from point 1

Where are they used? Artificial neural networks and problems in need of solution

In theory, the number of types of problem that can be dealt with using ANNs is limitless, as the methodology is broadly applicable, and the problems spring up as fast as the questions that society, for whatever reason, poses. So, let us remind ourselves of the criteria that must be satisfied for the adoption of an ANN to be worthwhile:

- The problem is a complex one
- Other methods have not produced satisfactory results
- An approximate solution to the problem is valuable, not only a certain or best solution
- An acceptable solution to the problem offers great savings in human and/or economic terms
- There exists a large case history demonstrating the 'strange' behavior to which the problem pertains

Fig. 6



Schematic comparison of artificial neural network (ANN) with other analysis techniques. Compared with other analysis techniques, ANNs are useful when one has a problem with a lot of available data but no good theory to explain them.

Figure 6 summarizes the conditions that best claim for neural networks analysis.

A special feature of neural networks analysis: variables selection

ANNs are able to simultaneously handle a very high number of variables notwithstanding their underlying nonlinearity. This represents a tremendous advantage in comparison with classical statistics models in a situation in which the quantity of available information is enormously increased and nonlinearity dominates. With ANNs one is concerned neither about the actual number of variables nor about their nature. Owing to their particular mathematic infrastructure, ANNs have no limits in handling the increasing amounts of variables that constitute the input vector for the recursive algorithms.

Discriminant analysis, logistic regression, or other linear or semilinear techniques typically employ a limited number of variables in building up their model: those with a prominent linear correlation with the dependent variable. In other words, for them a specific criterion exists, for example a correlation index, which indicates which of the variables available has to be used to build a predictive model addressed to the particular problem under evaluation.

ANNs, being a sort of universal approximation system, are able to use a wider range of information available, and also variables with a very poor linear correlation index. For this reason, the natural approach is to include all of the variables that, according to clinician experience, might have an *a priori* connection with the problem being studied. When the ANNs are exposed to all these variables of the data set, they can very easily approximate all the information available during the training phase. This is the strength, but unfortunately also the weakness, of ANNs.

In fact, almost inevitably, a number of variables that do not contain specific information pertinent to the problem being examined are processed. These variables, inserted in the model, act as a sort of 'white noise' and interfere with the generalization ability of the network in the testing phase. When this occurs, ANNs lose some potential external validity in the testing phase, with consequent reduction in the overall accuracy. This explains why, in absence of systems able to select appropriately the variables containing the pertinent information, the performance of ANNs might remain below expectations, being only slightly better than classical statistical procedures.

One of the most sophisticated and useful approaches to overcoming this limitation is to define the subgroup of variables to be selected with another family of artificial adaptive systems: evolutionary algorithms (EAs).

Evolutionary algorithms

At variance with neural networks, which are adaptive systems able to discover the optimal hidden rules explaining a certain data set, EAs are artificial adaptive systems able to find optimal data when fixed rules or constraints have to be respected. They are, in other words, optimization tools, which become fundamental when the space of possible states in a dynamic system tends toward infinity. This is just the case of variables selection. Given a certain, large amount of dichotomous variables (for example 100), the problem of defining the most appropriate subgroup to best solve the problem under examination has a very large number of possible states and exactly: 2^{100}. The computational time required to sort all possible variables subsets to submit them for ANN processing would be in the order of million years; a so-called nonpolynomial hard mathematical problem.

The introduction of variable selection systems generally results in a dramatic improvement in ANN performance.

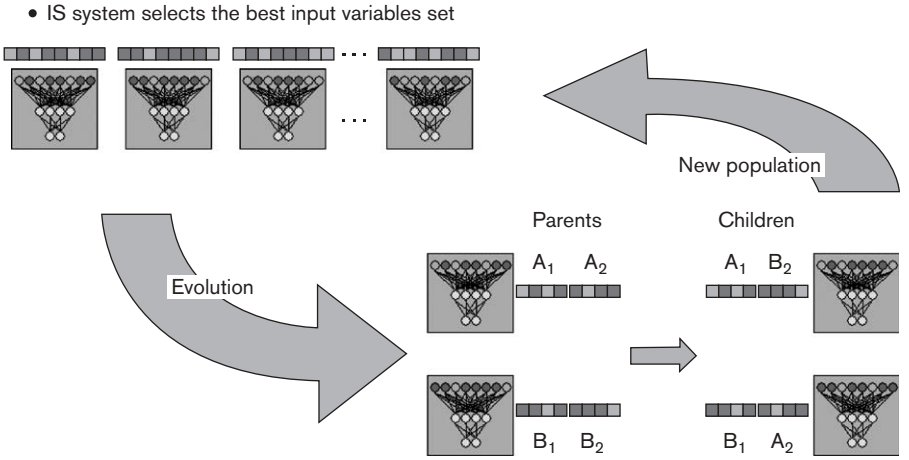
Input selection (IS) is an example of the adaptation of an EA to this problem.

This is a selection mechanism of the variables of a fixed dataset, based on the EA GenD [26] (Fig. 7). The IS system becomes operative on a population of ANNs, each of them capable of extracting a different pool of independent variables. Through the GenD EA, the different 'hypotheses' of variable selection, generated by each ANN, change over time, generation after generation. When the EA no longer improves, the process stops, and the best hypothesis of the input variables is selected and employed on the testing subset. The goodness-of-fit rule of GenD promotes, at each generation, the best testing performance of the ANN model with the minimal number of inputs.

An example of an application of IS system is described in the paper by Buscema *et al.* [27], which also contains a theoretical description of the neuroevolutionary systems.

As is shown in Table 1, which refers to the papers in which variables reduction has been performed, the application of ANNs to a selected subset of variables has systematically yielded to an improvement of predictive capacity. These results demonstrated that deep mining of such highly nonlinear data could extract enough information to construct a near optimum

Fig. 7



Input selection (IS). Each individual of the population codes a subset of different independent variables. The best solution along the generation process consists in the smallest number of independent variables able to reach up the best classification rate in testing phase.

Table 1 Effect of variables reduction in enhancing the classification performance of ANNs

Papers	No. of variables	Overall accuracy (%)	No. of variables	Overall accuracy E (%)
Andriulli *et al.* (a) [28]	45	64.5	34	79.6
Andriulli *et al.* (b) [28]	45	69.0	31	88.6
Pagano *et al.* [29]	24	74.4	9	89.4
Sato *et al.* [30]	199	75.0	60	83.5
Pace *et al.* [31]	101	77.0	44	100
Lanher *et al.* [32]	37	96.6	30	98.4

ANN, artificial neural network.

classifier/predictor able to provide effective support for the physician's decision, after an appropriate data processing

Discussion

As can be seen from the general literature, and from the papers after this introduction in this special issue of *European Journal of Gastroenterolgy and Hepatology*, ANNs constitute new, potent, computational tools able to outperform the diagnostic and prognostic accuracy of classical statistical methods.

In complex gastrointestinal disorders such as, for example, dyspeptic syndrome, chronic pancreatitis or nonerosive reflux oesophagitis (NERD), many symptoms at presentation in an uninvestigated patient might lie at the crossroads of a group of disorders. For instance, they could be an expression of functional disorders as well as of organic diseases. By definition, complex chronic diseases have heterogeneous origins in which various mechanisms participate to a different extent in different patients. Consequently, techniques belonging to classical statistics, such as discriminant analysis, which assume linear functions underlying key pathogenetic factors, normal or quasi normal value distribution, and reduced contribution of outliers through average computation, might incorrectly represent the complex dynamics of sociodemographic, clinical, genetic, and environmental features that may interact in these patients. ANNs, to the contrary, take advantage of modern mathematical theories inspired by the life sciences and seem to be able to extract from the data information that is not considered useful by traditional approaches.

The IS operated by the EAs deserves a specific comment. The potential impact of this particular subset of variables could be high in routine practice, reducing substantially the number of questions about present symptoms and the number of clinical and laboratory data collected, with obvious advantages from a logistic and economic point of view.

Therefore, from a medical perspective, this review has clearly demonstrated that available data, unusable by conventional assessment methods, need not be considered as being 'useless'. Such data can be, and actually is precious, providing, after appropriate processing, helpful information and effective support for the doctor's and the patient's decisions.

We have also seen that fine results were reached in the optimization phase using more costly, complex systems

composed of neural networks and evolutionary subsystems working together as hybrid organisms. Such protocols require a consistent level of expertise. Their use is recommended only when the input–output relationships are sufficiently complex, as in the examples in this review. When this criterion is not met it is simpler and less expensive to use traditional methods, such as discriminant analysis.

A second important feature is represented by the improved transfer of evidence derived from clinical research to single patient level.

Both physicians and their patients are put under pressure by the potential, as well as actual, risk of future disease and by the uncertainty, and the associated anxiety, of anticipating the outcome and treatment of a disease. The increasing demand of individualized treatment, of specific diagnoses suited to a single subject, and of accurate prognosis by modeling risk factors in the context of that particular individual requires a new age of statistics, the statistics of the individual.

Physicians are more and more aware of the fact that the individual patient is not an average representative of the population. Rather she/he is a person with unique characteristics, predicaments, and types and levels of sensibility; they are also aware that prevention (primary, secondary, tertiary) can be effective on the population but not necessarily for the targeted individual patient, and that reaching the objective of a prevention guideline based on a specific protocol does not necessarily mean a favourable outcome in that given individual. Clinical epidemiology and medical statistics are not particularly suited to answering specific questions at the individual level. They have, after all, been developed primarily to focus on groups of individuals and not on single individuals.

We know that any kind of statistical inference loses its meaning in the absence of a 'sample', which by definition requires a number greater than 1. For this reason, predictive models can fail dramatically when applied to the single individual.

When applied to a single individual, the degree of confidence for a model that on average has an accuracy rate of 80% at a group level, can drop substantially.

Suppose that a predictive model for diagnostic or prognostic classification has been developed and validated in a study data set, and that it allows an overall accuracy of 80%. Suppose that the confidence interval of this predictive rate is 10% (75–85%). We now assess a group of new individuals with our tool. We can reasonably expect to make classification mistakes in the order of 15–25%. In other words 15–25, out of 100 new patients would be misclassified. If I am a new patient and I have been classified in a certain way (e.g. diseased), I might think that I have an 80% chance of being correctly classified (75% in the worst and 85% in the best case). Unfortunately, the confidence interval of this classification at my level would not be equal to that of the group as, in the case of misclassification, I would suffer from an all-or-nothing situation (correct diagnosis vs. incorrect diagnosis). This would mean a 100% difference. In other words, at single participant level the confidence interval would be wider than the mean accuracy rate at a group level.

As is not possible to transform the single individual into a group of individuals on which one performs some statistics, one could do the opposite; that is, treat a single individual with a group of statistics. In other words, this means using several independent classification models, which make different errors while retaining a similar average predictive capacity, on the same individual. ANNs allow this.

It is theoretically possible to train hundreds of neural networks with the same data set, resulting in a sizeable assembly of ANNs with a similar average performance but with the predisposition to make different mistakes at individual level owing to the fact that they differ slightly in their architecture, topology, learning laws, starting weights of interconnections between nodes, order of presentation of cases during training cycle, number of training cycles, and so on.

In this way, it is possible to produce a large set of neural networks with high training variability able to process independently a set of new patients to predict a diagnostic category, a survival plausibility, or a drug dosage level. For each patient, several hundreds of answers would be generated. Therefore, when a new patient has to be classified, thanks to this sort of 'parliament of independent judges' acting simultaneously, a specific nonparametric distribution of output values could be obtained with resulting descriptive statistics (mean, mode, median, measures of variance, confidence interval, etc.). It is interesting to note that the classification output of neural networks is generally expressed according a fuzzy logic scheme, along a continuous scale of 'degree of membership' to the target class, ranging from 0 (minimum degree of membership) to 1 (maximum degree of membership). According to the above reasoning, it could be possible to establish a suitable degree of confidence of a specific classification in the individual/patient, overcoming the dogma, which excludes the possibility of making statistical inference when a sample is composed by just one individual.

A word of caution: these theories must be proved in the real world. As with medications, these systems also need to have their effectiveness studied in routine medical settings. We need to demonstrate that the use of application software based on trained neural networks can increase the quality of medical care delivery and/or reduce costs. We feel optimistic, and using the words of D. Hollander [33] 'Given the rapid advances in computer hardware, it is likely that sophisticated ANN programming could be made available to clinics and outpatient care facilities and could save time and resources by guiding the screening clinician towards the most rapid and appropriate diagnosis and therapy of patients with gastrointestinal problems.'

Acknowledgements

The authors want to express their gratitude to Alessandra Mancini for her precious help in reviewing the literature and assisting in manuscript preparation.

Conflict of interest: none declared.

Annotated references

● Of special interest
●● Of outstanding interest

1 McCulloch WS, Pitts WH. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 1943; **5**:115–133.

●● The first mathematical model of logical functioning of brain cortex (formal neuron) is exposed in the famous work of McCulloch and Pitts.

2 McClelland JL, Rumelhart DE, editors. *Explorations in parallel distributed processing*. Cambridge, Massachusetts: MIT Press; 1986.

●● This book is the historical reference text on the neurocomputing origins, which contains a comprehensive compilation of neural network theories and research.

3 Anderson JD, Rosenfeld E, editors. *Neurocomputing: foundations of research*. Cambridge, Massachusetts: MIT Press; 1988.

● An interesting book that collects the most effective works on the development of neural networks theory.

4 Hebb DO. *The organization of behavior*. New York: Wiley; 1949.

●● In this landmark book is developed the concept of the 'cell assembly' and explained how the strengthening of synapses might be a mechanism of learning.

5 Marr D. Approaches to biological information processing. *Science* 1975; **190**:875–876.

● In this article, Marr, writing about his theoretical studies on neural networks, expanded such original hypotheses.

6 Rosenblatt F. The Perceptron. A probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958; **65**:386–408.

●● The first neural network learning by its own errors is developed in this hystorical work.

7 Widrow G, Hoff ME. Adaptive switching circuits. Institute of radio engineers, Western Electronic show & Convention, Convention record. 1960, part 4:96–104.

8 Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, editors. *Parallel distributed processing*. Vol. I. Boston: MIT Press; 1986. pp. 318–362.

9 Personnaz L, Guyon I, Dreyfus G. Collective computational properties of neural networks: new learning mechanisms. *Phys Rev A* 1986; **34**: 4217–4228.

10 Gallant SI. Perceptron-based learning algorithms. *IEEE Transaction on Neural Networks* 1990; **1**:179–192.

● An important paper that describes the main learning laws for training the neural networks models.

11 Wasserman PD. *Neural computing: theory and practice*. New York: Van Nostrand; 1989.

12 Aleksander I, Morton H. *An introduction to neural computing*. London: Chapman & Hall; 1990.

Two books containing systematic expositions of several neural networks models.

13 Fahlman SE. *An empirical study of learning speed in back-propagation networks technical report CMU-CS-88-162*. Pittsburg: Carnegie-Mellon University; 1988.

14 Le Cun Y. Generalization and network design strategie. In: Pfeifer R, Schreter Z, Fogelman-Soulie F, Steels L, editors. *Connectionism in perspective*. North Holland: Amsterdam; 1989. pp. 143–156.

● The performances and possible generalizations of Back-Propagation Algorithm are described by Fahlman and Le Cun.

15 Hinton GE. How neural networks learn from experience. *Sci Am* 1992; **267**:144–151.

In this article there is a brief and more accessible introduction to Connectionism.

16 Matsumoto G. Neurocomputing. Neurons as microcomputers. *Future Gen comp* 1988; **4**:39–51.

● The Matsumoto article is a concise and interesting review on Neural Networks.

17 NeuralWare. *Neural computing*. Pittsburgh, Pennsylvania: NeuralWare Inc.; 1993.

18 CLEMENTINE user manual. Integral Solutions Limited; 1997.

19 Von der Malsburg C. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* 1973; **14**:85–100.

20 Willshaw DJ, Von der Malsburg C. How patterned neural connection can be set up by Self-Organization. *Proc R Soc London B* 1976; **94**: 431–445.

● The early network model that performs self-organization processes has been exposed in papers from Von der Malsburg and Willshaw.

21 Kohonen T. *Self-organization and associative memories*. Berlin-Heidelberg-New York: Springer; 1984.

22 Kohonen T. The self-organizing map. *Proceedings IEEE* 1990; **78**: 1464–1480.

●● The most well-known and simplest self-organizing network model has been proposed by T. Kohonen.

23 Carpenter GA, Grossberg S. The ART of adaptive pattern recognition by a self-organizing neural network. *Computer* 1988; **21**:77–88.

24 Carpenter GA, Grossberg S. A massively parallel architecture for a self-organizing neural pattern recognition machine. In: Carpenter GA, Grossberg S, editors. *Pattern recognition by self-organizing neural networks*. Cambridge, MA: MIT Press; 1991.

● These works of Grossberg and Carpenter are very interesting contributions regarding the competitive learning paradigm.

25 Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput* 1998; **7**:1895–1924.

●● This article describes one of the most popular validation protocol, the 5×2 cross validation.

26 Buscema M. Genetic doping algorithm (GenD): theory and applications. *Exp Syst* 2004; **21**:63–79.

● A seminal paper on the theory of evolutionary algorithms.

27 Buscema M, Grossi E, Intraligi M, Garbagna N, Andriulli A, Breda M. An optimized experimental protocol based on neuro-evolutionary algorithms: application to the classification of dyspeptic patients and to the prediction of the effectiveness of their treatment. *Artificial Intelligence Med* 2005; **34**:279–305.

●● A complex work that used techniques based on advanced neuro/evolutionary systems (NESs) such as Genetic Doping Algorithm (GenD), input selection (IS) and training and testing (T&T) systems to perform the discrimination between functional and organic dyspepsia and also the prediction of the outcome in dyspeptic patients subjected to *Helicobacter pylori* eradication therapy.

28 Andriulli A, Grossi E, Buscema M, Festa V, Intraligi M, Dominici PR, *et al*. Contribution of artificial neural networks to the classification and treatment of patients with uninvestigated dyspepsia. *Digest Liver Dis* 2003; **35**: 222–231.

● A paper assessing the efficacy of neural networks to perform the diagnosis of gastro-oesophageal reflux disease (GORD). The highest predictive ANN's performance reached an accuracy of 100% in identifying the correct diagnosis; this kind of data processing technique seems to be a promising approach for developing non-invasive diagnostic methods in patients suffering of GORD symptoms.

29 Pagano N, Buscema M, Grossi E, Intraligi M, Massini G, Salacone P, *et al*. Artificial neural networks for the prediction of diabetes mellitus occurrence in patients affected by chronic pancreatitis. *J Pancreas* 2004; **5** (**Suppl 5**):405–453.

● In this work several research protocols based on supervised neural networks are used to identify the variables related to diabetes mellitus in patients affected by chronic pancreatitis and presence of diabetes was predicted with an accuracy higher than 92% in single patients with this disease.

30 Sato F, Shimada Y, Selaru FM, Shibata D, Maeda M, Watanabe G, *et al*. Prediction of survival in patients with esophageal carcinoma using artificial neural networks. *Cancer* 2005; **103**:1596–1605.

●● This study is the first to apply the ANNs as prognostic tools in patients with esophageal carcinoma using clinical and pathologic data. The ANN models demonstrated an high accuracy for 1-year and 5-years survival prediction and their performance was superior when compared to corresponding Linear Discriminant Analysis models.

31 Pace F, Buscema M, Dominici P, Intraligi M, Grossi E, Baldi F, *et al*. Artificial neural networks are able to recognise GERD patients on the basis of clinical data solely. *Eur J Gastroenterol Hepatol* 2005; **17**:605–610.

● In this prelimimary work, data from a group of patients presenting with typical symptoms of gastro-oesophaceal reflux disease (GORD) and diagnosed using oesophagoscopy and pH-metry were processed by different ANN models. The network with the highest predictive performance achieved an accuracy of 100% in identifying correct diagnosis (positive or negative GORD patients) whereas the traditional discriminanr analysis obtained an accuracy of 78%.

32 Lahner E, Grossi E, Intraligi M, Buscema M, Delle Fave G, Annibale B. Possible contribution of advanced statistical methods (artifical neural networks and linear discriminant analysis) in the recognition of patients with suspected atrophic body gastritis. *World J Gastroenterol* 2005; **11**:5867–5873.

● A study that suggest the use of advanced statistical methods, such as ANNs but also LDA, to better address bioptic sampling during gastroscopy in patients with suspected atrophic body gastritis.

33 Hollander D. Is artificial intelligence an intelligent choice for gastroenterologists? *Digest Liver Dis* 2003; **35**:212–214.